Sex differences in performance in Double Award Science GCSE

John F Bell
University of Cambridge Local Examinations Syndicate
1 Hills Road, Cambridge, CB1 2EU

Paper presented at the British Educational Research Association Annual Conference (September 11-14 1997: University of York)

Abstract

In this paper sex differences in science performance of year 11 pupils will be considered. This paper represents a progress report of ongoing work and is not intended to be complete and the conclusions are only tentative.

For many years, researchers have been concerned with the sex difference in performance in favour of males on physics-based tasks and the differential uptake by females at school, in higher education and science-based careers. Double award science was introduced to improve scientific literacy, to reduce sex imbalances in take-up, and to encourage greater uptake of science in further and higher education.

By considering the 1996 results for Nuffield Co-ordinated Sciences: Double Award 1772, this paper will consider how combined sciences has influenced sex differences. In particular, performance differences in the physical and biological sciences will be discussed.

Introduction

In the last twenty years, there have been many changes that have influenced the science education of year 11 school children in England. These changes have included the change from CSE and O-level examinations to the GCSE, a large change in the numbers of year 11 pupils and the introduction of the National Curriculum.

Before the introduction of the national curriculum, a major concern was the Biology/Physics polarisation in the option choices of boys and girls and the resulting limitation of later career opportunities (DES, 1980, Royal Society and Institute of Physics, 1982; Kelly, 1982; Kahle, 1985; Johnson and Murphy, 1986).

However, support for balanced science courses was not universal. When balanced integrated science courses were introduced, some teachers were concerned. Reid and Ryles (1989) surveyed the attitudes of science teachers. They found that two thirds of teachers were concerned that A level studies would become relatively more difficult and about the reduction in the amount of science available to the most able fifteen- and sixteen-year-olds. They also questioned whether it would have a positive impact on the science education of girls. They were concerned that there would be a further deterioration in girls' attitudes in response to a diet of 'force-fed science'. It was suggested that the reluctance of girls to go on to study science was not going to be overcome by a syllabus which provided more science with unchanged teaching methods, rather than different science with better teaching methods.

With the introduction of the National Curriculum, almost all pupils in state schools were required to take science and the apparent performance differences had reversed. Elwood (1995) noted that gender-related differences in performance in public examinations in the UK have attracted press interest, e.g., *Independent* (1993, 1994), the *Sunday Times* (1994), the *Times Educational Supplement* (1994). Arnot, David and Weiner (1996) claimed that:

'There has been a notable move towards gender equality in Science and Mathematics at GCSE, which in 1985, showed a clear gender gap in performance in favour of male students. Changing examining procedures in Science may partly be the reason, although the separate sciences of Physics, Chemistry and Biology, though much depleted in entry, also exhibit similar declines in expected male performance.'

Elwood (1995) reviewed the performance in GCSE and noted that

> ...the patterns of performance in the GCSE are not unique, but are part of wider trends that support the assumption that the old stereotypes of girls' and boys' performance in examinations and assessment systems no longer hold firm.

This lead to headlines such as 'They're falling rapidly behind girls at school. Are boys in terminal decline?' (Independent, 1994).

This conclusion is surprising given the history of science education. For many years, researchers have been concerned with the sex difference in performance in favour of males on physics-based tasks and the differential uptake in science of females at school, in higher education and in science-based careers (Casserly, 1975; DES; 1980, Johnson and Murphy, 1984, 1986; Kelly, 1981). For example, in the Assessment of Performance Unit's science surveys, on average, boys performed better than girls on physics questions at ages 11 (year 6), 13 (year 8) and 15 (year 11 pupils who were tested in November, when most would still be only 15-years-old) (Russell et al., 1988; Schofield et al., 1988; and Archenhold et al., 1988). The size of the difference in performance increased between year 8 and year 11 as a result of sex differences in science subject uptake (Johnson and Bell, 1987). This was one of the reasons that integrated double science GCSE was introduced. This change, combined with the National Curriculum, has greatly reduced sex differences in science uptake (Daniels and Bell, 1990; Bell, in press). By 1995 almost all pupils in state schools were required to take science and the differences in gender have apparently reversed. Smithers and Robinson (1994) concluded that double science had met two of its objectives: to improve scientific literacy and to correct gender imbalances. However, they found that a third objective of greater take-up of sciences at A-level had not been met. In 1994, 5.8% of the males and 1.7% of the females in the cohort of seventeen-year-olds took A level physics examinations. 1.2% of males and 0.3% of females obtained a grade A (figures derived from DfEE, 1995). The three-subject structure of A levels, poor teaching in primary school, and the labour market were suggested as reasons for those who do double award science at GCSE not going on to study science at A level.

This paper will demonstrate that, by considering the examination results in greater detail, the sex differences still exist. It will investigate the sex differences in science performance by investigating the performance on the components of GCSE Nuffield science: Co-ordinated Sciences Double Award (1772). In 1996, assessment was by means of course work for the Attainment Target Sc1: Scientific Investigations, and terminal written examinations for Sc2: Life and Living Processes, Sc3: Materials and their Properties, and Sc4: Physical Processes. There were three tiers of entry for written examinations. Each tier contained questions set on specific levels as indicated in Table 1. This meant that there was a total of nine written examinations for this syllabus.

Table 1: Tier structure for GCSE Nuffield Co-ordinated Sciences: Double Award 1772

Tier of entry	Questions set on levels	Grades Available
Basic	3, 4, 5, 6	G-D (C)
Central	5, 6, 7, 8	F-B (A)
Further	7, 8, 9 10	D-A*

(The award of grade C for the Basic tier and Grade A for the central tier was considered in exceptional cases only)

In this preliminary analysis, the differences in performance on parts of questions between the sexes are discussed. Various classifications of the question parts are considered and are used to suggest further research directions.

Overall GCSE results

In table 1, the cumulative grade distributions of the science subjects are presented. These have been expressed as a percentage of the total cohort which gives a clearer idea of the level of performance in the population as a whole. For example, 4.8% of year 11 boys in 1994 obtained a grade B or better in GCSE Biology in 1994. Obviously many pupils did not enter GCSE Biology and could get a grade B in it if they were able, but presenting the data this way avoids misleading comparisons of sex differences. In all three separate sciences, more boys than girls obtained each grade. Although not directly illustrated in the table, the pupils of either sex tended to obtain better results than pupils entered for Science Double Award. Candidates entered for Science Single Award usually obtained less than a grade C. These differences are caused by selection effects: low attaining candidates tend to be entered for the single Award and not the separate sciences.

Table 2: Cumulative grade distribution of 15-year-olds by sex for science subjects in 1994
(% of pupil in cohort obtaining grade or better)

Subject	Sex	A*	A	B	C	D	E	F	G
Biology	M	0.5	2.4	4.8	6.9	7.8	8.2	8.3	8.4
	F	0.4	1.9	3.5	4.8	5.3	5.6	5.7	5.8
Chemistry	M	0.7	2.7	5.0	6.9	7.8	8.3	8.4	8.5
	F	0.5	1.8	3.2	4.2	4.7	5.0	5.1	5.0
Physics	M	1.1	3.2	5.2	7.2	8.1	8.5	8.7	8.7
	F	0.5	1.6	2.7	3.8	4.3	4.6	4.7	4.7
Science Single Award	M	0.0	0.1	0.6	1.4	3.9	6.5	8.4	9.3
	F	0.0	0.2	1.1	2.3	5.5	8.3	10.3	11.2
Science Double Award	M	2.1	7.0	19.9	33.2	49.3	60.4	66.7	68.8
	F	2.1	7.4	20.9	34.4	50.6	62.3	69.0	71.4
Other Sciences	M	0.1	0.2	0.4	0.8	1.1	1.4	1.7	1.8
	F	0.0	0.1	0.3	0.5	0.7	0.8	0.9	0.9

(Source: DfE (1995)

The differing levels of the attainments of the entries and the differential uptake for the science examination make interpretation of the sex differences in the results difficult to interpret. To satisfy the National Curriculum requirements, there are three choices: either all three separate sciences, Science Double Award or Science Single Award. Since no candidates were entered for both separate subjects and the Science Double Award, it is possible to make a rough estimate of the sex differences (including Science Single Award would make interpretation difficult). Combining the results for Biology and Science double award gives 9.4% of boys and 9.3% of girls obtaining a grade A. For Chemistry and Science Double Award, the percentages are 9.7 for boys and 9.3 for girls, and for Physics and Science Double Award, the percentages are 10.2 and 9.0 respectively. The full results are presented in Table 2.

Table 3: Cumulative grade distribution of 15-year-olds by sex for science subjects combinations
(% of pupil in cohort obtaining grade or better in either subject assuming no candidate was entered for both)

Subject	Sex	A*	A	B	C	D	E	F	G
Biology & Science	M	2.6	9.4	24.7	40.1	57.1	68.6	75.0	77.2
Double Award	F	2.5	9.3	24.4	39.2	55.9	67.9	74.7	77.2
Chemistry & Science	M	2.8	9.7	24.9	40.1	57.1	68.7	75.1	77.3
Double Award	F	2.6	9.2	24.2	38.6	55.3	67.3	74.1	76.4
Physics & Science	M	3.2	10.2	25.1	40.4	57.4	68.9	75.4	77.5
Double Award	F	2.6	9.0	23.6	38.2	54.9	66.9	73.7	76.1

These findings contradict the suggestion that the only subject in which boys are outperforming girls at GCSE is Mathematics. However, these differences are relatively small. This analysis makes the implicit assumption that candidates who obtained a grade A for Double Award Science would have obtained A grades for the single subjects.

Results for Nuffield Coordinated Sciences

The analysis in the previous section does not tell the whole story. To do this, it is necessary to consider examination performance in more detail. This is demonstrated by considering data from the 1996 Nuffield Coordinated Sciences Double Award GCSE.

Because of their origins in O-level syllabuses, the Nuffield Science syllabuses attracted a better than average entry compared with other double certificate GCSE science examinations. This is illustrated in Table 3 below. For example, 43% of candidates taking Nuffield examinations obtained a grade B or higher compared with 27% for the combined entry for all double award science syllabuses. There are no important sex differences in the grade distributions for the awarded grades for the Nuffield Syllabus and for double award syllabuses in general.

Table 4: Final grades by sex for Nuffield Co-ordinated Sciences Double Award 1996 and for all Double Awards 1996
(Cumulative percentage distributions)

	Final		Results for all double awards	
GRADE	Female	Male	Female	Male
A*	5	6	4	4
A	18	18	11	10
B	42	43	29	28
C	62	62	50	49
D	82	82	71	71
E	94	93	87	88
F	98	97	96	96
G	99	98	99	99
U	99	99	100	100
X	100	100	-	-
Total	13969	13957	461971	462181

(X denotes absent candidates)

In Table 4, the grade distribution by attainment target is presented. This was achieved by combining the component grades for each tier by each Attainment Target for the written examination papers. The grade A* is not awarded for individual components. There is a clear pattern of sex differences. There are sex differences by attainment target. For example, 4% more female candidates obtain a grade C or above for the component assessing Sc2: Life and living processes, and 11% more male candidates obtain a grade C or above for the component assessing Sc4: Physical Processes. There is also a large difference in favour of female candidates on the component assessing Sc1: Investigations. Sc1 is assessed by coursework.

Table 5: Grade distributions by attainment target for Nuffield Double Award Science
(June 1996 data. Cumulative percentages)

	Sc1: Investigations		Sc2: Life		Sc3: Matter		Sc4: Physical	
Grade	Female	Male	Female	Male	Female	Male	Female	Male
A	41	35	20	17	17	17	14	20
B	69	63	38	35	35	33	30	41
C	85	80	60	56	56	56	49	60
D	92	88	80	77	77	77	67	77
E	96	95	92	91	92	92	82	89
F	98	97	98	97	98	97	93	95
G	99	98	99	98	99	98	95	96
U	100	100	99	99	99	99	99	99
X	100	100	100	100	100	100	100	100
Total	13969	13957	13969	13957	13969	13957	13969	13957

The method of deriving overall examination grades specified in the mandatory code of practice for awarding GCSEs (SCAA, 1996) means that a poor performance on one component can be compensated for by a stronger performance on another. To investigate the effect of this compensation, the difference between the component grade and the overall grade for the whole examination was investigated. There is a number of interesting features. The component grade on Sc1 that is assessed as coursework tends to be the same as or higher than the overall grade and the examination-based component grades tend to be lower than or the same as the overall grade. Secondly, there are large sex differences. For Sc1, 63% of females and 52% of males obtained higher grades on this component than on the whole examination. For Sc2, just under one quarter of the female candidates and approximately one third of the male candidates performed worse on the component than they did on the examination. The sex difference is relatively small for Sc3. The largest difference is for Sc4; more than half the female candidates obtain a lower grade on the component compared with the examination overall. This is true for only one quarter of the male candidates. The differences between components are described in greater detail in Bell (1997).

Before considering the question structure it is instructive to consider the grade boundaries for these papers. There are given in Table 5 below. These boundaries indicate that the awarding committee considered that the Sc2 paper was more accessible than the other two. There is no requirement that the boundary marks for the different components should be same although there would be an intention to set papers for which the boundary marks would be similar.

Table 5: Grade boundaries for the Component of the further tier components of Nuffield Double Award Science

Boundary	Sc2: Life	Sc3: Materials	Sc4: Physical
A/B	71	63	63
B/C	60	50	50
C/D	49	39	39
D/U	38	29	27

The results of the above analysis and the boundary generate the following questions. Why is the Sc 2 examination easier than the other two examinations and has enough allowance been made for the relative difficulty of the Sc 4 examination?

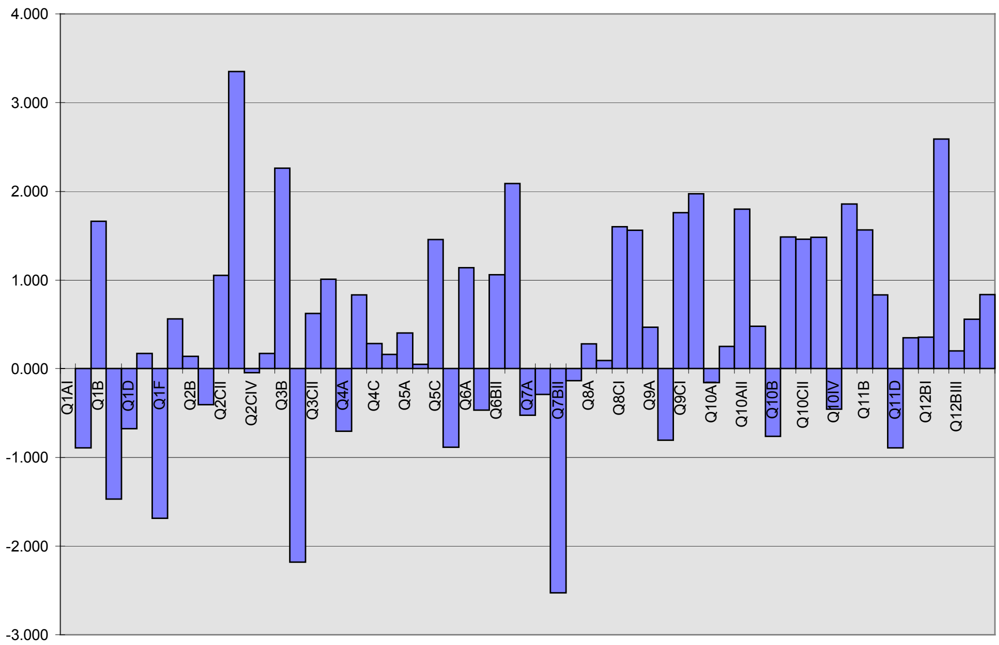
Analysis of Individual Question Parts

In this section, the performance of candidates on the individual parts of the science questions in the further tier of the 1996 Nuffield Science Double examination will be considered. It is intended to extend the analysis to the all the tiers of Double Award, the Single Award and the separate subjects. There are common questions across the differing examinations within a tier and across adjacent tiers and across the Single Award and the sperarate subjects..

Random samples of scripts were drawn for each component and the data at the individual question part level was entered. Usually only the total mark on a component is entered and used in the administration of the examination. For each question a short descriptive synopsis is written. For each question part, a t-test was carried out by sex to identify those questions for which the difference was significant at the 5% level. Because of the large number of t-tests carried out some of the significant differences will have occurred by chance. Because the entry for this syllabus is not a random sample, there is a need for caution in interpreting the differences between the sexes. The important feature of the analysis is the relative differences. It is not possible to state that, in general, male candidates performed better than female candidates on physics questions only that this is true for candidates entered for the further tier of Nuffield Double Award science. This is further complicated by the fact that this analysis is restricted to candidates entered for the further tier. These candidates would be expected to perform reasonably well on all components.

In Figures 1, 2 and 3, the t-statistics for each question part in examination order are presented for papers 7, 8 and 9 respectively. For Sc2 (Paper 7), girls outperform boys on most of the question parts but the differences are rarely statistically significant. The boys obtain better marks on a few questions.

Figure 1: Bar chart of t-statistics for each question part for paper 7: Sc2 Life and Living processes
(positive difference in favour of female candidates. Questions in examination order)



The average performance for four question parts is significantly better for female candidates than male question parts and for two question parts the reverse is true. The synopses for these question parts are as follows:

Female candidates better

Q2CIII	Identify the phenotypes of two people in a family tree
Q3B	Explain why not all cells of a fruit fly have the same number of chromosomes
Q6BIII	Suggest an explanation for colour of offspring's ears from a pair of experimental rabbits
Q12BII	Give two differences between two graphs of nutrient uptake

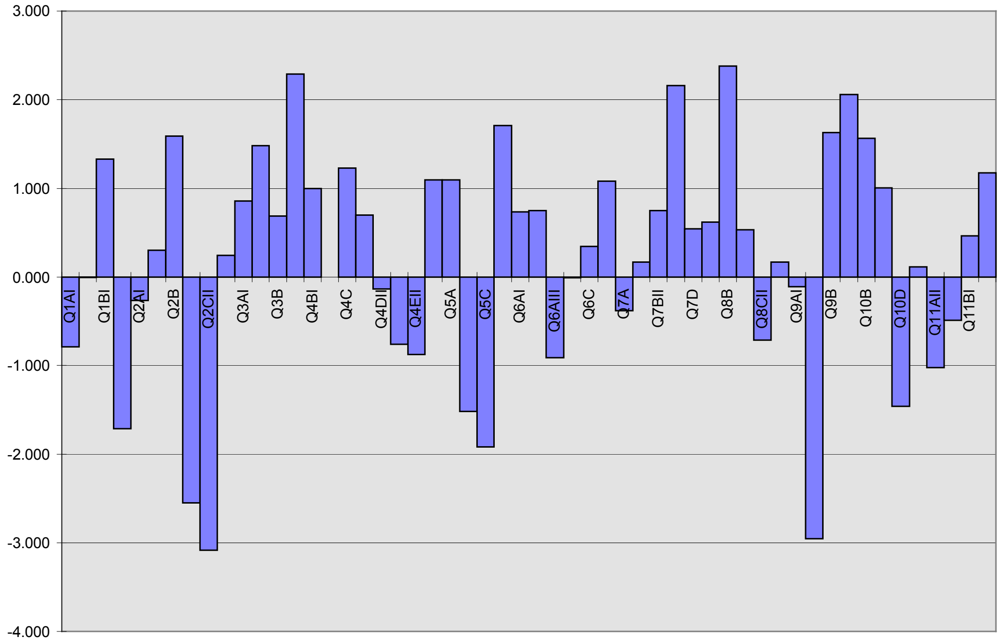
Male candidates better

Q3CI	Identify what faulty nuclear division of chromosomes is called
Q7BII	Explain two ways that cutting down trees caused desertification

There is no obvious pattern that explains these differences.

In Figure 2, a bar chart of the t-statistics of the question parts for the further tier SC 3 examination are given. The mean scores for girl were higher than the mean scores for boys for a substantial proportion of the question parts. However, for most of the question parts the difference was not significantly different.

Figure 2: Bar chart of t-statistics for each question part for paper 8: Sc3 (positive difference in favour of female candidates. Questions in examination order



The mean score for female candidates was significantly better on the following question parts:

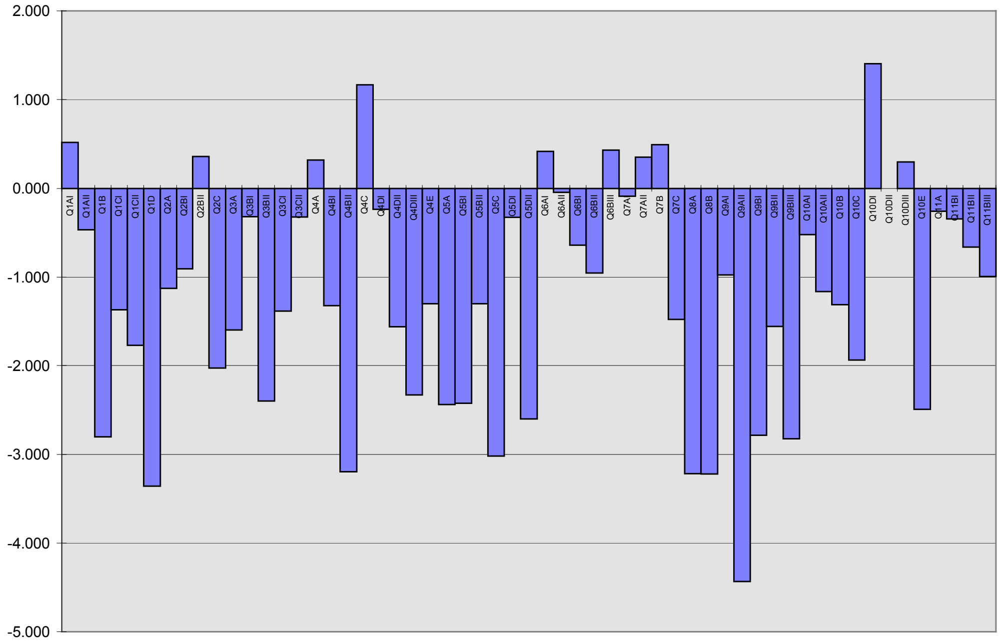
Q4A Complete the sentence 'The elements in the Periodic table are arranged in order of their ...'
Q7C Using information in a table calculate the total energy transfer when two compounds react
Q8B Complete the equation for the change of aluminium ions to atoms
Q10A What is meant by Isotopes?

and the mean score for male candidates was significantly better for

Q2CI From a set of graphical formulae of five hydrocarbons choose the solid petroleum jelly
Q2CII From a set of graphical formulae of five hydrocarbons choose the gas
Q9AII Use the nature of layers of earth crust to describe the speed of shock waves

The bar chart of t-statistics (Figure 3) is dramatically different from the other two. On average male candidates performed better than female candidates on the majority of questions.

Figure 3: Bar chart of t-statistics for each question part for paper 9: Sc4 (positive difference in favour of female candidates. Questions in examination order)



For the Sc 4 examination, there were seventeen question parts for which there was a significant sex difference in performance in favour of male candidates and no sex differences in favour of female candidates. The synopses of the question parts are given below:

Q1B	Explain why a switch in an RCD does not move when the current is the same
Q1D	Explain why a fuse is better than an RCD
Q2C	What is the maximum speed of the car on the big dipper?
Q3BII	Explain why the theoretical temperature rise is too high
Q4BII	Suggest one other factor explaining the acceleration of a rocket
Q4DIII	What is the change in velocity after burning the fuel?
Q5A	Explain why a bulb in an electric circuit is dim
Q5BI	Identify a type of transformer
Q5C	Explain why the transformer only works with AC
Q5DII	What assumption has been made in the calculation of voltage?
Q8A	Calculate the force needed to operate a lever
Q8B	Explain why the weight of the pole means the lever needs less force
Q9AII	Which building has the lowest natural frequency?
Q9BI	Relate natural frequency to the frequency of an earthquake vibration
Q9BIII	Explain why some buildings do not sway in an earthquake
Q10C	Use a graph to estimate the speeds of four galaxies
Q10E	Estimate the size of the universe using data from the table

It is interesting that many of the questions require an explanation of physical phenomena and that relatively few involve calculations. The next stage of the analysis was to classify the questions by various methods in order to identify the features that are associated with sex differences.

There are many ways of classifying questions. For the purposes of this paper three classifications will be considered:

Question Type,
Response Type,
Mathematics Content.

In an effort to interpret the pattern of differences the APU assessment framework was considered. Because this reflected a view of science that encompassed both the procedural aspects of science and the applications of conceptual knowledge, this needed to be modified. The APU question bank did not include questions that expected the simple recall of facts. For the purpose of this study, an additional category has been added. Because these papers only consider the written tests only written categories of the APU framework are considered.

The new category *Recalling scientific information* includes questions in which candidates are asked to give the standard name for some scientific feature, or a standard meaning of some standard terms or an explanation of a common theory. The answers to these questions only require remembering an answer and only require a minimal understanding of the information. This type of question was not included in the APU surveys because before the introduction of the National Curriculum pupils followed many different science course combinations.

All questions require some minimal conceptualisation of the phenomena under discussion. Questions in the *Applying Science Concepts* are therefore based on the apparent necessity to use (not the mere possibility of using) ideas that a pupil is unlikely to have met outside science lessons. The category *Interpreting presented information* includes questions that test the procedural skills involved in science but do not include questions that require the recall of science concepts (which are placed in the category *Applying Science Concepts).* Also included in this category are questions in which all the information necessary to answer them is given in the question. Because questions in the first category *Use of graphical and symbolic representation* were rarely used in these examinations, they were placed in the category *Interpreting presented information.* The categories are listed in Table 6.

Table 6: Question Categories

ID	Category	Description
U	Use of graphical and symbolic representation	Reading and inserting in tables, charts and graphs. Constructing Tables and Charts and graphs. Using scientific symbols and conventions e.g. chemical equations, circuit diagrams, and section drawings.
I	Interpreting presented information	Either question in which all the information need to answer them is presented or do not require ideas met in science lessons
A	Applying Science Concepts	Explicitly require an understanding of science concepts
R	Recalling scientific information	Requires the use of memory to recall the name of process or feature or the meaning of a name

The category *Applying Science Concepts* includes questions from the attainment targets Sc2, Sc3 and Sc4. These questions explicitly require an understanding of taught science concepts. They require the application of such ideas in situations representative both of normal school science and of more wide ranging settings.

For some questions, it is difficult to determine which of the three categories should be applied. For the purposes of analysis the categorisation of the questions was carried out for parts of questions for which a mark was explicitly awarded. There are two features to note. For whole examination questions, all three categories could be used. In addition, inspecting the mark schemes reveals that within question parts marks could be awarded to different categories (for example, in *Applying Science Concepts* a mark may be awarded for recalling the concept and further marks for the actual application). However, without remarking the scripts, it is not possible to separate these implicit parts.

There are some questions where it is debatablee whether the knowledge used is common sense or a science concept. This was particular the case for Biology. In addition there is a problem deciding, for some question parts whether it is possible that the candidates had memorised an answer or whether they had applied a concept.

Although there is a clear relationship between this classification and Bloom's taxonomy (Bloom, et al., 1956; Nitko, 1983), *Recalling Scientific Information* would be defined as knowledge. There is an important difference. The categories used here allow for the fact that knowledge, comprehension and application interact. The problem with using Bloom's taxonomy is that it is designed to classify educational objectives and the cognitive demand of questions (or parts of questions).

The proposed relationship between difficulty and question categories is presented in Figure 4. It is assumed that there are two dimensions relating to the question: the amount of skills required to answer the question and the amount of knowledge required to answer the question. *Recalling scientific information* requires only low levels of scientific procedural skills but varying amounts of knowledge and questions in this category would be placed along the bottom of the figure. *Interpreting presented information* requires low levels of knowledge and should be placed along the left of the figure. *Applying Science Concepts* requires both skills and knowledge and should be placed in the central area of the figure. The design of the figure emphasises the fact that the questions cannot be ordered in difficulty by category. It also illustrates the expected performance of differing types of candidates: an intelligent but lazy candidate would tend to have high levels of skills but would have low levels of knowledge from lack of revision so would be able to answer questions on the left side of the figure in the *Interpreting presented information* category. An less intelligent but diligent candidate would tend to be able to answer questions along the bottom of the figure in the *Recalling Scientific Information* category.

Figure 4: Relationship between difficulty and question categories

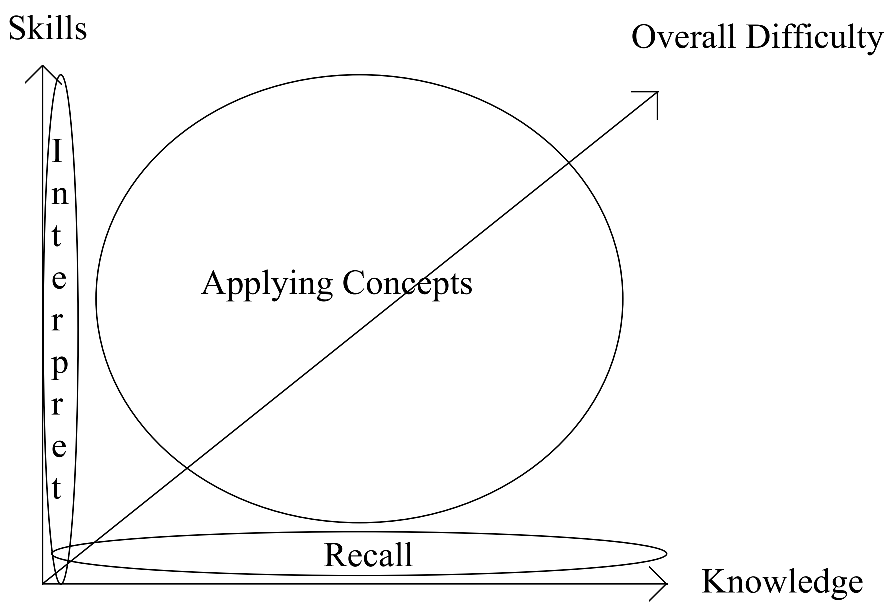


Figure 4 also emphasises the fact that the overall difficulty of questions may vary within the categories. An *Applying Science Concepts* question may be easier than a *Recalling Scientific Information* question when the level of recall and skill is lower. The questions used in an examination could result in any order of difficulty for the three categories.

In the table below, the number of marks awarded to the different types of question are given. The number of marks indicate the weight given to each activity in the examination. Note that this different from the number of questions.

Table 7: Distribution of question marks by categories for each examination

Category	Sc2	Sc3	Sc4
Recalling Scientific Information	40	45	49
Interpreting presented Information	33	32	6
Applying Science Concepts	32	28	47

There is one major difference in the table; there are more questions in Sc4 that require the application of science concepts. This is the result of the National Curriculum. In the syllabus document, it states that

"Questions related to statements of attainment which begin 'Understand quantitative relationships...' (mainly levels 7 and 8 in Sc4) will require candidates to recall formulae. For questions relating to Statements which begin 'be able to use...' it is expected that candidates will be able to recall formulae. Formulae and equations not given in Statements will be provided on the question paper.

The list of formulae and relationships which candidates will be asked to recall is given as appendix D." (20 physics formulae and relationships are listed in this appendix).

The effect of this was to force more questions into the category *Applying Science Concepts.* The candidates were never explicitly asked to write down a formula. They were usually expected to recognise that it was needed, recall it and carry out a calculation with it.

Response Type

An attempt was also made to categorise the response types. These are listed in Table 8.

Table 8: Response types for science question parts

ID	Response Type	Description
M	Multiple Choice	Question for which candidates can select an answer from a given set of options
A	Non-written	Answer does not require the use of words. Usually this means a number
D	Diagram	The response is a drawing
S	Single word/phrase	E.g. naming a process
W	Written response	Answer requires one or two sentences
E	Extended written	Answer may require more than two sentences

Although the examination is not a multiple choice test, there are some questions for which it is possible to guess the answer. These have been classified as having a multiple choice response type.

An extended written response was one in which either several blank lines were given for the response on the examination paper or several marks awarded for a written response. This classification does not indicate an essay but rather a paragraph in the context of this examination. Some questions are effectively multiple choice because candidates can pick an answer from a list given. The distribution of question marks by response type is given in Table 9. Because some response types tend to have more marks per question part (e.g. extended response) than others (e.g. single word/phrase) it does not reflect the number of questions.

Table 9: Distribution of question marks by response type for each examination
(total number of marks awarded by question category)

Response type	Sc2	Sc3	Sc4
Multiple Choice	6	10	1
Non-written	14	32	41
Diagram	0	0	2
Single word/phrase	11	5	6
Written response	62	42	47
Extended Written response	12	16	5

The number of marks for written responses was highest for Sc2 and lower for Sc3 and Sc4. The reverse pattern applied for non-written responses.

Mathematics Content

One possible source of difficulty is the amount of mathematics required by the question. This was investigated using a simple classification of the question parts. The levels of mathematics content are listed in Table 10.

Table 10: Levels of Mathematics Content for science question parts

ID	Mathematics Content	Description
N	None	No mathematics at all.
W	Whole numbers	Simple arithmetic involving whole numbers.
S	Single stage	More complex arithmetic which could involve the use of a calculator but only one operation e.g. substituting numbers in an equation and calculating the values.
M	Multi-stage	As above but requires more algebraic manipulation or more than one calculation.
O	Other mathematics	Usually involving the creation and interpretation of graphs.

The distribution of marks by mathematics content is given in Table 11. The level of mathematics content increases from Sc2 to Sc4. In Sc2 the only mathematics is interpreting tables and charts. In Sc3 the mathematics content tends to be a mixture of simple whole number sums and more complex arithmetic and in Sc4 the mathematics content tends to be more complex arithmetic.

Table 11: Distribution of question marks by mathematics content for each examination (total number of marks awarded by question category)

Mathematics Content	Sc2	Sc3	Sc4
None	87	73	60
Whole Numbers	0	13	0
Single Stage	0	14	37
Multi-stage	0	5	3
Other mathematics	18	0	2

Obviously the three ways of classifying questions are not independent. *Recalling Scientific Information* questions have no mathematics content and an extended writing response is unlikely to be expected for question involving a complex calculation.

Analysis of classification by performance

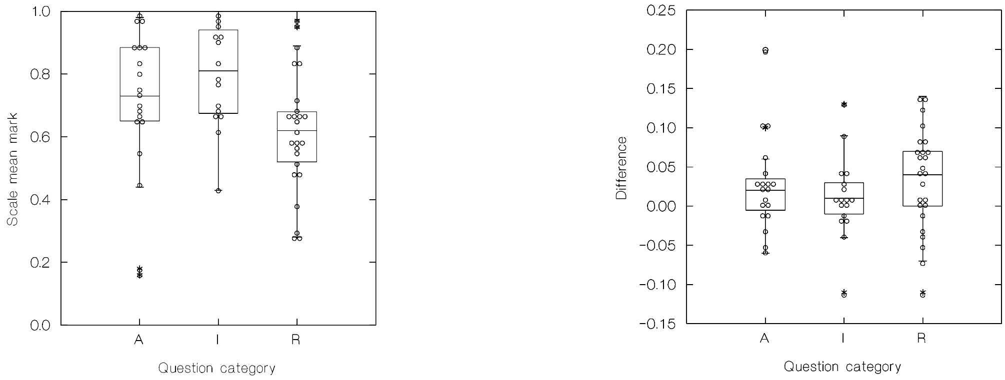
For the purposes of these analyses, all question parts were scaled on to a 0-1 scale and a scaled mean mark was calculated for each question part. The analysis described here is exploratory and intended to suggest future analysis and uses graphical techniques rather than formal statistical models. It should be recognised also that this analysis is restricted to samples of scripts from the further tier examination.

Sc2: Life and Living Processes

The mean scaled mark by each category is presented as box plots in figure 5. A dot plot has been superimposed on each box plot. Each circle on the dot plot represents a question part. When question parts have the same mean score (within the resolution of the plot) they are placed side by side. Two sets of box plots have been presented. The first is the scaled mean score which ranges from 0 to 1 for all question parts. This is a measure of the difficulty of the question. The second is the difference between the female candidates' mean scores and the male candidates' mean scores. A negative value indicates that males candidates performed better on a question part compared with female candidates. From figure 5, clearly the question parts categorised as *Interpreting presented information* tend to be easy and that some question parts categorised as *Recalling scientific information* are the most difficult on the paper. There is also a tendency for female candidates to perform better on question parts categorised as *Recalling scientific information.*

Figure 5: Box plots by question category for Sc2

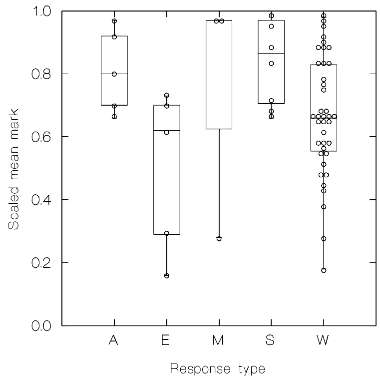
(a) Overall scaled mean **(b) Mean differences**



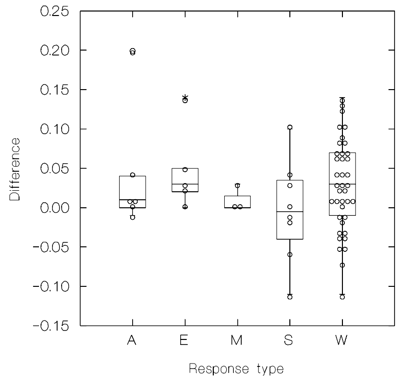
Most of the question parts on Sc2 were classified as having a written response. This is illustrated by the number of dots for some types in figure 6. It is noticeable, however, that responses requiring a single word proved to be relatively easy. However, the classification by response type is not particularly useful for this examination.

Figure 6: Box plots by response type for Sc2

(a) Overall scaled mean



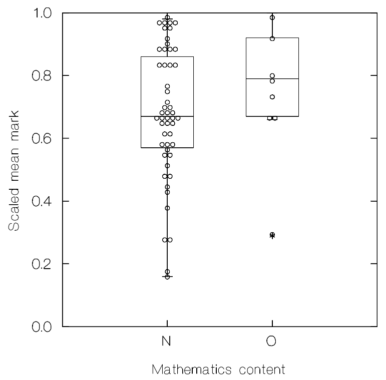
(b) Mean differences



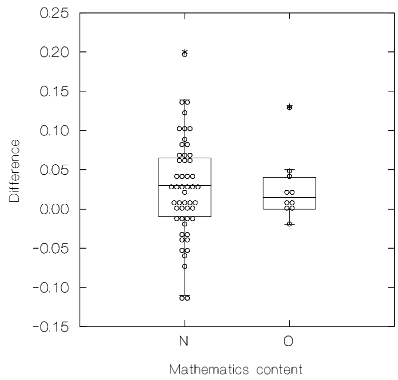
Because of the lack of mathematics content for this attainment target, figure 7 is not particularly useful. There is no evidence that the small number of question parts relating to graphs and tables were either easier or more difficult than other question parts in the examination.

Figure 7: Box plots of mathematics content for Sc2

(a) Overall scaled means



(b) Differences



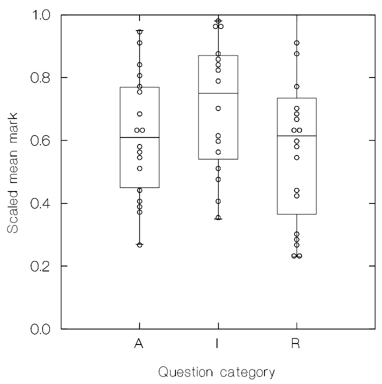
In summary for Sc2, only question category provided any useful information. *Recalling Scientific Information* is found to be the most difficult by candidates but female candidates tend to outperform males on this category.

Sc3: Materials and their properties

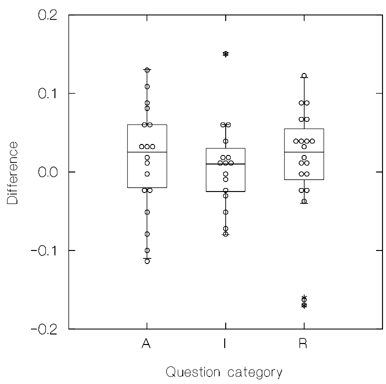
From figure 8, clearly candidates tended to find question parts categorised as *Interpreting Presented Information* as the easiest. Again there was some evidence to suggest that female candidates may be better at *Recalling Scientific Information*.

Figure 8: Box plots by question category for Sc3

(a) Overall scaled mean



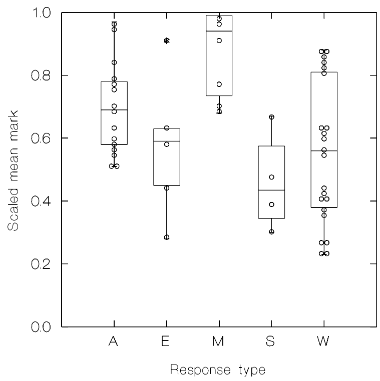
(b) Mean Differences



The response type classification is more useful for Sc3 compared with Sc2, with some question types having enough questions to suggest the possibility of differences. The questions that involved guessing (type M) proved to be the easiest, followed by the questions that did not require a written response. For the sex differences, the figure suggests that female candidates tend to perform better on question parts involving written responses.

Figure 9: Box plots by response type for Sc3

(a) Overall scaled means



(b) Mean Differences

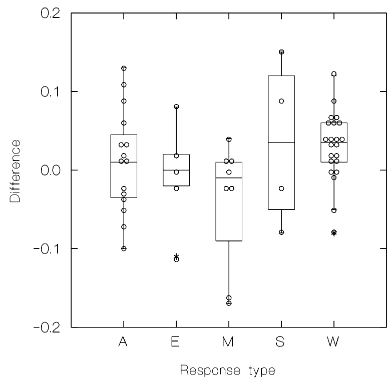
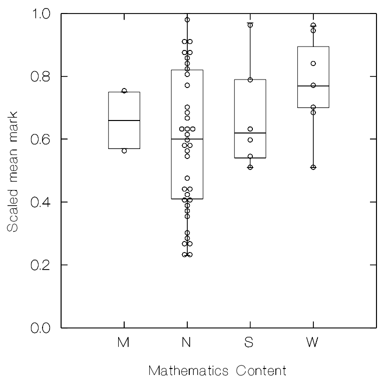


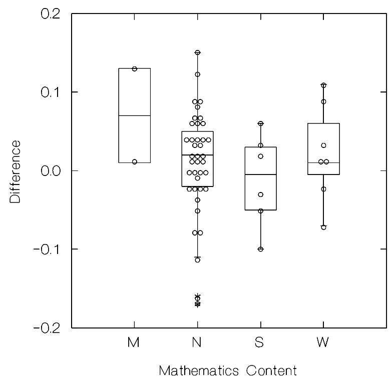
Figure 10 suggests that the relatively small number of question involving whole number calculations tended to be the easiest on the paper but did not uniformly favour either sex.

Figure 10: Box plots by mathematics content for Sc3

(a) Overall Scaled means



(b) Mean Differences

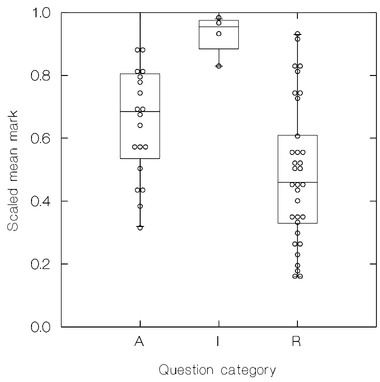


Sc4: Physical Processes

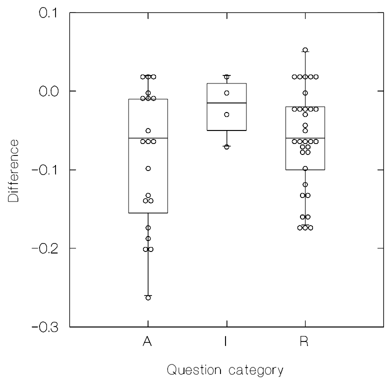
For Sc4, only the categories *Applying Scientific Concepts* and *Recalling Scientific Information* were represented with reasonable numbers of questions. Question parts in the *Recalling Scientific Information* category were more likely to be difficult than question parts in the *Applying Scientific Concepts* category. However, the sex differences in favour of male candidates tend to be more pronounced for the category *Applying Scientific Concepts.*

Figure 11: Box plots by question category for Sc4

(a) Overall scaled means



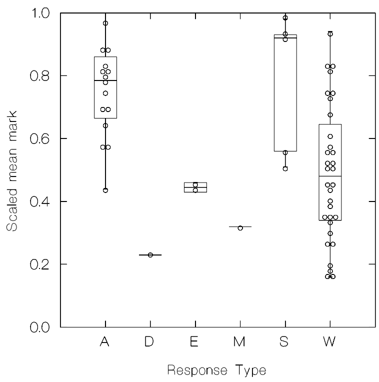
(b) Mean Differences



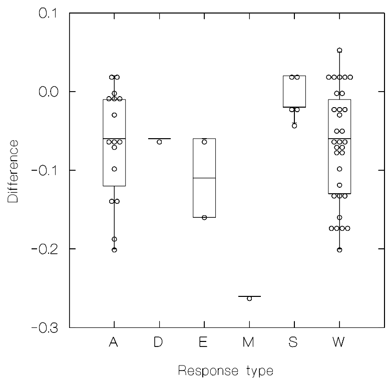
Only two response types, non-written (A) and written (W), are represented in sufficient numbers for this attainment target. From Figure 11, clearly questions parts with a written response type tend to be more difficult. However, they do not explain the sex differences.

Figure 12: Box plots by response type for Sc4

(a) Overall scaled means

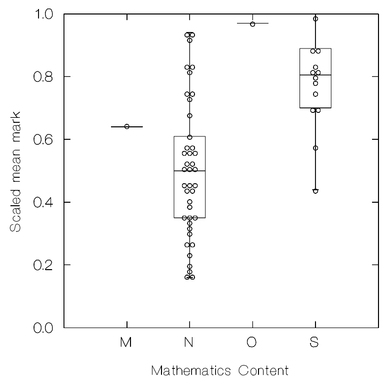


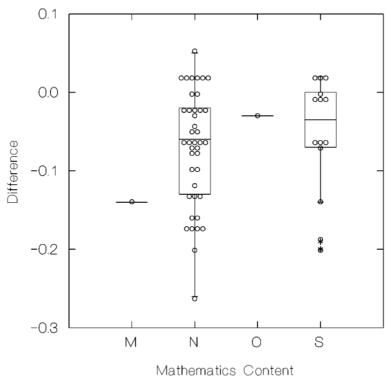
(b) Mean Differences



From figure 12, questions with no mathematics content tend to be the most difficult (they would, of course, include all the questions in the category *Recalling Scientific Information*).

Figure 13: Box plots by mathematics content for Sc4





The results from Sc4 are interesting. The mathematics content is not the source of difficulty but rather an inability to recall scientific information for the able pupils considered here. Does this imply that in recent years there has been too much emphasis on process skills and not enough on background knowledge? In addition, the sex difference in Physics performance seems to be related to poor performance *Applying scientific concepts.* This weakness in conceptual understanding could explain why so few continue with physical science. The explanations for this will be considered in the discussion.

Discussion

It is important to recognise that this preliminary analysis is restricted to candidates entered for the further tier which means that if they have been entered for the appropriate tier they will represent the most able part of the cohort. This could explain why these candidates found that, on average, the category *Interpreting Presented Information* was the easiest for this group. An able candidate who has not revised sufficiently would be able to answer question in this category but would not be able to answer questions which require *Recalling Scientific Information* or *Applying Scientific Concepts.* .

One of the findings of this paper is that most candidates found the physics (Sc 4) paper harder than the papers for the other two sciences. The relative ease of biology (Sc 2) for this tier might be explained by the high proportion of questions that have been classified as *Interpreting Presented Information* and by the lack of mathematical content. It is noticeable that a greater proportion of the questions on the biology paper could be answered by reference to experiences outside the classroom (e.g. television news) and by application of reasoning skills. The low proportion of *Interpreting Presented Information* questions in Physics may explain the difficulty for this examination. However, the reason for this low proportion is strongly influenced by the fact that candidates are required to recall twenty simple formulae which some might consider as not a particular heavy demand.

The analysis in this paper indicates that the sex gap in physics performance has been masked by the introduction of double science GCSE. The results presented in this paper suggest a series of issues that should be considered:

The Physics Gap still exists
The interpretation of the published figures has often been incorrect. The traditional pattern of sex differences still applies in science at this age, although the compulsion to follow balanced science courses has reduced their magnitude.

The unevenness of the grade profiles
Is it appropriate to report science achievement by awarding the same grade twice as required by SCAA? The compensation that results from the combining of component marks may also go some way in explaining the view that demanding material is not covered satisfactorily. The components have a

series of assessment objectives. This means that the knowledge, skills and understanding specified by the attainment targets only account for 50% of the marks.

Analyses of the Adequacy of the Double Award GCSE as preparation for A-level
For example, a grade A in GCSE physics almost certainly means that a candidate received an A for Sc4. This is not necessarily the case for candidates with a grade A double science GCSE. This could include candidates with grade B on Sc4 which has been compensated for by superior performance on the other components. It would, therefore, be expected that grade A double science GCSE candidates would, on average, obtain worse results at A-level than grade A GCSE physics candidates.

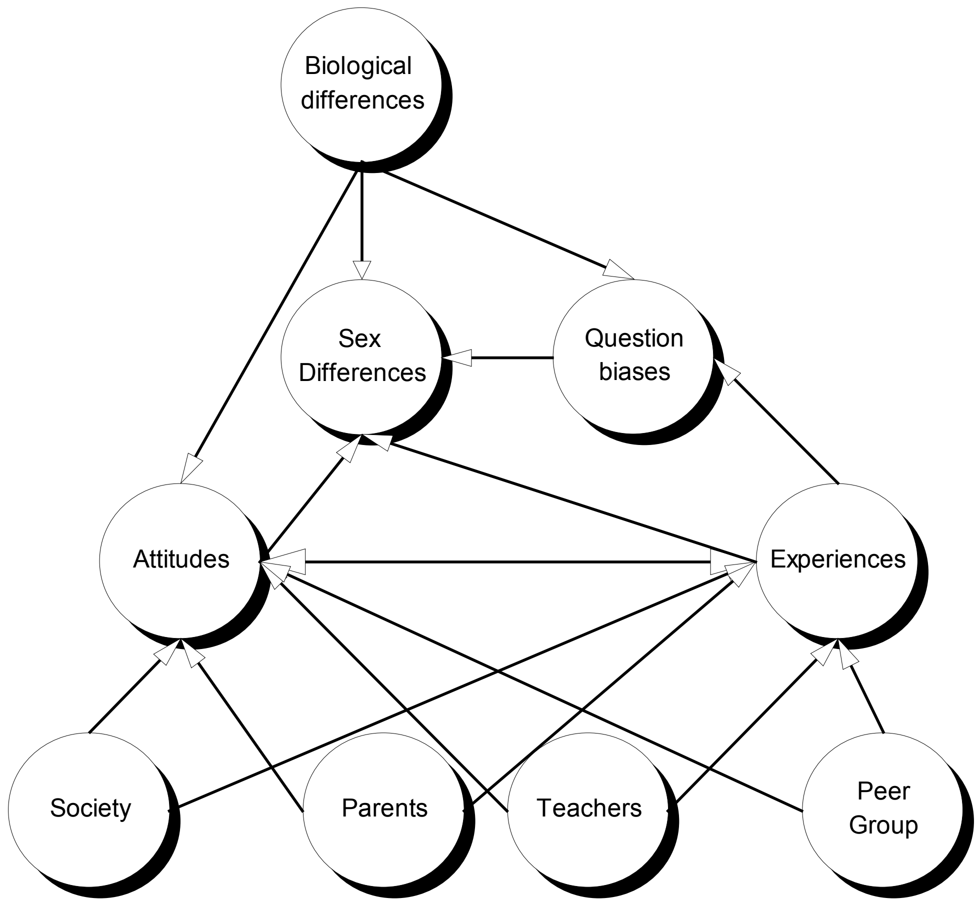
Compulsion to study Physics is not enough to close the Physics gap
There were many underlying causes of the continued sex differences in science uptake and performance in further and higher education. There has been a number of factors which have been proposed as reasons for this difference. These include attitudes to future careers, out-of-school interests, biological differences, the nature of science and the nature of science teaching.

Further research is needed into the weaker performance on certain types of question by female candidates.
Within UCLES the work on the Question Difficulty Project which is experimenting with alternate forms of the same questions may provide an insight into the causes of these sex differences.

Poor performance on questions requiring recall
For all three attainment targets, the question parts which were the most difficult came from the category *Recalling Scientific Information*. Other researchers have noted that when science processes are separated from science knowledge, gender differences occur only on knowledge (Hueftle, Rakow and Welch, 1983; and Zimmerer and Bennett, 1987). These authors also found that when both are combined, males outperform females especially in physical sciences. This finding is not supported by the analysis in this paper.

There is still a need to research into the causes of these sex differences. These reasons have been set out in the concept map given as Figure 14. This attempts to illustrate the relationships between various proposed causes of the sex differences. In this diagram, question biases is taken to mean sex differences resulting from question format and context and not the physics content of the question. Note that experiences and attitudes are linked by a two way arrow: bad experiences can result in bad attitudes and bad attitudes can lead to poor experiences.

Figure 14: Concept map illustrating possible causes for sex differences in physics performance



There has been a considerable amount of research into the causes placed at the bottom of the diagram. Some of these relate to attitudes to science in the community. One of the major factors affecting this choice is the perceived usefulness for jobs or further education (Johnson and Bell, 1987, Schofield, et al., 1988, Kelly, 1981; Price and Talbot, 1982; Nash *et al.,* 1984; CDSS, 1985, Taber, 1989). Other studies have suggested that girls tend to have less confidence in their academic abilities and their potential to undertake professional careers (Davies and Meighan, 1975; Furlong, 1986; Russell et al., 1989). Girls and boys have also been shown to aspire to different jobs (Bould and Hopson, 1983; Davis, 1987), and some fathers may well feel that their work is not suitable for their daughters (Davis, 1987). In research undertaken by the Engineering Council in 1991, five-year-old children were asked to assess jobs and activities according to those suitable for men, those suitable for women and those suitable for both genders. They thought that science was more suitable for men than, for example, firefighting or climbing mountains. Girls are likely to seek jobs in fields where they already know women workers (Furlong, 1986). Another feature of girl's reactions to the society's traditional stereotypes is the feeling that they must conform or appear less feminine (Delamont, 1980).

As a result of peer group pressure, some girls may also fear the consequences of academic success and deliberately avoid it (Faulkner, 1991). For example, a team of eight ten-year-old boys and girls were asked at a science investigation event organised by the Association for Science Education to mount an inquiry. The boys went away with the equipment to do the enquiry and left the girls to prepare a poster to report the team's result. When asked, the children said they were simply doing what each of them was "good at". It has also been hypothesised that this gap might be attributed to the very different leisure activities of young boys and girls out of school (Johnson and Murphy, 1986). The activities preferred by boys were of a kind offering greater opportunities to develop practical skills and acquire an appropriate grounding for later conceptual learning in physics. There is no reason to expect that changes in education have had large effects on these patterns of behaviour. Browne and Ross (1991) concluded from observations of a large number of pre-school children that from a very young age children develop clear ideas about what girls do and what boys do. The activities girls choose to take part in more than boys were labelled by Browne and Ross as *creative*. Boys on the other hand opted for *constructional* activities.

In APU practical investigation tasks, although girls were found to be just as competent as boys, more girls than boys lacked confidence in handling equipment. The effect has been linked to observations of boys' and girls' behaviours in laboratories where boys typically dominate the use of apparatus and equipment (Whyte, 1986).

Other areas of concern are the nature of science and the aggressive culture of experimental research (Kelly, 1985; Keller, 1985) and the interactions between teachers and pupils in science lessons (Crossman, 1987). Bentley and Watts (1988) make a case for the radical revision of science provision for girls (and for boys). For example, Versey (1990) recommended positive steps that all science teachers could take to enhance access and motivation of pupils.

There is evidence of a decline in positive attitudes toward science as pupils progress through school (Doherty and Dawe, 1985; Piburn and Baker, 1993). A factor that emerged clearly as a negative influence on attitude was the growing abstraction and complexity of science classes. While this applies to both sexes, it was found to be more marked in the case of girls. Extensive research (Kelly, 196; Stables and Harvey, 1986; Parker and Rennie, 1986; Levin et al., 1991; Solomon and Harrison, 1991) has been carried out into the difference between boys' and girls' attitudes towards science. The work highlights the underachievement of girls in the physical sciences as well as their less positive attitude, and suggests strategies for improving girls' attitude towards this branch of science.

There is the suggestion that girls' cognitive capacity with respect to physical science is not equivalent to that of boys and that this reinforces the prejudicial practice of directing girls away from science (Casserly, 1975). For example, Gray (1981) concluded that nothing should be done about girls and science, arguing that 'we should rather celebrate the differences between the sexes as a contribution to the diversity society requires.' These arguments are undermined by the simple fact that assessment data show that *most* girls are as competent as *most* boys, i.e. there are very large overlaps in the performance curves for girls and boys in science achievement tests. A fuller discussion of biological differences can be found in Geary's (1996) paper in *Behavioral and Brain Sciences* and the commentary by various researchers that follows it.

With all these factors influencing the scientific education of females, the results of this paper are not surprising. Clearly it takes more than a requirement to take balanced science and carefully written examinations and syllabuses to affect the sex differences in science education. Various strategies have been suggested to improve the situation.

There has been research into gender stereotyping. In a recent review of subjects in British schools, Archer (1992) concluded that there is modest evidence that gender stereotyping is diminishing. Archer and Macrae (1991) found that there was less stereotyping evident in their study of ratings of the masculinity-femininity of different subjects than in one published a decade earlier (Weinreich-Haste, 1981), but many subjects still had gender stereotypes.

Koballa (1992) argues that most of the science related attitudes held by teachers and pupils are acquired incidentally rather than as a result of planned effort, and if science educators wish to change these attitudes they would find it beneficial to study contemporary persuasion theory. The research of Crawley and Koballa (1994) supports the use of belief-based persuasive messages to promote positive attitudes to studying chemistry. In a study comparing the attitudes of science and non-science college students towards science material, Goglin and Swartz (1992) determined that non-science students have high anxiety levels and low levels of motivation towards science. Woolnough (1997) noted that many school physics courses were considered by the majority of pupils to be difficult, dull, theoretical, impersonal and requiring hard work. However, he also noted that able pupils, particularly those likely to continue with science, considered these courses to be easy and stimulating.

Holloway (1997) argued that women have a different management style from men. They organise their laboratories in a less hierarchical way than men and prefer to work collaboratively rather than in competition. They are also more likely to be interested in scientific problems if they have a social relevance and could produce a social benefit.

This paper demontrate that sex differences still exist in science at year 11 and there is is need for turther research in this area.

References

ARCHENHOLD, W.F., BELL, J., DONNELLY, J., JOHNSON, S. and WELFORD, G. (1988) *Science at Age 15: A review of APU findings 1980-1984.* London: HMSO.
ARCHER, J. (1992) Gender stereotyping of school subjects, *The Psychologist*, 5, 66-69.
ARCHER, J. and FREEDMAN, S. (1989) Gender stereotypic perceptions of academic disciplines, *British Journal of Educational Psychology, 59, 306-313.*
ARCHER, J. and MACRAE, M. (1991) Gender perceptions of school subjects among 10-11 year olds, *British Journal of Educational Psychology, 61, 99-103.*
BANU, D.P. (1986) Secondary school student attitudes towards science. *Research in Science and Technological Education,* 4, 2, 195-202.
BELL, J.F. (1990) The curriculum of English 15-year-old pupils in 1984: was there a common core of subjects? *British Educational Research Journal*, 16, 1, 41-52.
BELL, J.F. (1991) Science performance and uptake by 15-year-old pupils in Northern Ireland. *Educational Research*, 33, 2, 93-102.
BELL, J.F. (1997). Evidence for the continued existence of the Physics 'gap' at age 16. *British Journal of Curriculum and Assessment,* 7, 35-39.
BELL, J.F. (in submission) Science uptake of year 11 pupils in 1984 and 1992. Submitted to *School Science Review*.
BENBOW, C.P. (1988) Sex differences in mathematical reasoning ability of intellectually talented preadolescents: their nature, effects, and possible causes. *Behavioural and Brain Sciences,* 11, 169-232.
BENTLEY, D. and WATTS, D.M. (1986) Courting the positive virtues: a case for feminist science. *European Journal of Science Education,* 8, 121-134.
BLACK, P. (1986) Integrated or co-ordinated science, *School Science Review*, 67, 241, 669-81.
BLOOM, B.S. (1976) *Human Characteristics and School Learning*. New York: McGraw Hill.
BLOOM, B.S., ENGELHART, M.D., FURST, E.J., HILL, W.H. and KRATHWOHL, D.R. (1956) *Taxomony of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain.* New York: David McKay and Co.
BONE, A. {1983) *Girls and Girls-Only Schools.* Manchester: EOC.
BOULD, J., and HOPSON, B. (1988) Sex differentiation and careers, in Marland, M. (ed.) *Sex differentiation and schooling.* London: Heinemann Educational.
BROWNE, N. and ROSS, C. (1991) Girls' stuff, boys' stuff: Young children talking and playing. In N. Browne (Ed.), *Science and Technology in the Early Years.* Milton Keynes: Open University Press.
BYRNE, E. (1978) *Equality of Education and Training for Girls (10-18 years).* Collection Studies: Education Series No. 9. Brussels: Commission for the European Communities.
CASSERLY, P.L. (1975) *An assessment of factors affecting female participation in advanced placement programs in mathematics, chemistry and physics.* Report to the National Science Foundation.
COLE, N.S. (1997) The ETS Gender Study: How Females and Males Perform in Educational Settings. Princeton, NJ: Educational Testing Service.
COLLEY, A., COMBER, C., and HARGREAVES, D.J. (1994) Gender effects in school subject preferences: a research note. *Educational Studies,* 20. 1, 13-18.
COMBER, L.C. and KEEVES, J.P. (1973) *Science education in nineteen countries.* Stockholm: Almqvist and Wiksell.
CRAWLEY, F.E., and KOBALLA Jr, T.R. (1994) Attitude Research in Science Education: Contemporary Models and Methods, *Science Education,* 78. 35-55.
CROSSMAN, M. (1987) 'Teachers' interactions with girls and boys in science lessons. In Kelly, A. (ed.) *The Missing Half: girls and science education.* Manchester: Manchester University Press.
DANIELS, S. and BELL, J.F. (1990) Balanced Science for all: provision for the sciences in secondary schools 1986-9. *School Science Review*, 71, 9-14.
DANIELS, S., and WELFORD, G. (1990) Self-assessment: pleasant surprise or harsh reality? *Evaluation and Research in Education*, 4, 1, 1-10.

DANIELS, S., and WELFORD, G. (1992) As they like it: Pupils' preferences, self-estimates and performance scores on science tasks. *Evaluation Research in Education.* 6, 1, 1-12
DAVIES, L. and MEIGHAN, R. (1975) A review of schooling and sex roles with particular reference to the experience of girls in secondary schools*, Educational Review,* 27, 3, 165-178.
DAVIS, D.J. (1987) Do you want your daughter or son in your trade? A study of the attitudes of job incumbents to females entering male dominated trades. *Comparative Education*, 23, 3, 279-285.
DEARING, R. (1996) *Review of Qualifications for 16-19 Year Olds*. *Mathematics and the Sciences.* London: SCAA.
DEE (1995) *Statistics of Education: Public Examinations GCSE and GCE in England*. London: HMSO.
DELAMONT, S. (1980) *The Sociology of Women.* London: George Allen and Unwin.
DELAMONT, S. (1994) Accentuating the positive: refocusing the research on girls and science. *Studies in Science Education, 23, 59-74.*
DES (1975) *Curricular Differences for boys and girls.* London: HMSO.
DES (1980) *Girls and Science.* HMI Matters for Discussion 13. London: HMSO.
DES (1982) *Science Education in Schools: A Consultative Document.* London: Department of Education and Science.
DES (1985) *Aspects of Secondary Education in England.* London: HMSO.
DOHERTY, J. and DAWE, J. (1988) The relationship between developmental maturity and attitude to school science: an exploratory study. *Educational Studies*, 11, 93-107.
DRIVER, R,. HEAD, J., and JOHNSON, S. (1984) The differential uptake of science in schools in England, Wales and Northern Ireland, *European Journal of Science Education*, 6, 19-29.
DRIVER, R., CHILD, D., GOTT, R., HEAD, J., JOHNSON, S., WORSLEY, C., and WYLIE, F. (1984) *Science in Schools. Age 15: Report No. 2.* London: APU.
DRIVER, R., GOTT, R., JOHNSON, S., WORSLEY, C., and WYLIE, F. (1982) *Science in Schools. Age 15: Report No. 1*. London: HMSO.
DWECK, C.S., DAVIDSON, W., NELSON, S., and ENNA, B. (1978) Sex differences in learned helplessness: The contingencies of evaluative feedback in the classroom. *Development Psychology,* 14, 268-278.
ELWOOD, J. (1995) Undermining gender stereotypes: examination and coursework performance in the UK at 16. *Assessment in Education*, 2, 283-303.
ERICKSON, G.L. and ERICKSON, L. (1984) Females and science achievement: evidence, explanations and implications. *Science Education,* 68, 63-89.
FAULKNER, J. (1991) Mixed-sex schooling and equal opportunity for girls: a contradiction in terms. *Research Papers in Education.* 6, 3, 197-223.
FRAZER, M.J. (1986) It's only a beginning. Science for all. Implications beyond 16. *School Science Review,* 68, 5-11.
FURLONG, A. (1986) Schools and the structure of female occupational aspirations. *British Journal of Sociology of Education,* 7, 4, 367-377.
GARDNER, P.L. (1987) Measuring ambivalence to science. *Journal of Research in Science Teaching*, 24, 241-247.
GEARY, D.C. (1996) Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences,*19, 2, 267.
GOGLIN, L., and SWARTZ, F. (1992) A quantitative and qualitative inquiry into the attitudes towards science of non-science college students. *Journal of Research in Science Teaching,* 29, 487-504.
GOTT, R., SCHOFIELD, B., DAVEY, A., GAMBLE, R., HEAD, J., KHALIGH, N., MURPHY, P., ORGEE, T. and WELFORD, G. (1985) *Science in Schools. Ages 13 and 15: Report No. 3.* London: APU.
GRAY, J.A. (1981) A biological basis for the sex differences in achievement in science? In A. Kelly (ed.) *The Missing Half: Girls and Science Education.* Manchester: Manchester University Press.
HANSFORD, B.C., and HATTIE, J.A. (1982) The relationship between self and achievement/performance measures. *Review of Educational Research,* 52, 123-42.
HAVEN, E.W. (1971) Factors associated with the selection of advanced academic mathematics courses in high school. *Dissertation Abstracts International,* 32: 1747A (University Microfilm No. 71-26027).
HOFFMAN, L. (1985) Differences in the subjective considerations of interest in physics and technology for boys ands girls. *Contributions to the Third GASAT Conference.* London: Chelsea.
HUEFTLE, S.J., RAKOW, S.J. and WELCH, W.W. (1983) Images of Science. Minneapolis: University of Minnesota, Science Assessment and Research Project.
INDEPENDENT (1993) Girls still outperform boys at GCSE. 18 June.

INDEPENDENT (1994) They're falling rapidly behind girls at school. Are boys in terminal decline? 18 October.
JOHNSON, S, (1987) Gender differences in science: parallels in interest, experience and performance, *International Journal of Science Education,* 9, 467-481.
JOHNSON, S. (1987) Early developed sex differences in mathematics and science in the UK, *Journal of Early Adolescence,* 7, 21-33.
JOHNSON, S. (1988) *National Assessment: the APU science approach.* London: APU.
JOHNSON, S. and BELL, J.F. (1987) Gender differences in science: option choices, *School Science Review,* 69, 247, 268-276.
JOHNSON, S. and MURPHY, P. (1986) *Girls and Physics. Reflections on APU survey findings.* APU Occasional Paper No. 4. London: Assessment of Performance Unit.
JOHNSON, S., and MURPHY, P. (1984) The underachievement of girls in physics: towards explanations, *European Journal of Science Education*, 6, 399-409.
JONES, A.T. and KIRK, C.M. (1990) Gender differences in students' interests in applications of school physics. *Physics Education*, 25, 308-313.
KAHLE, J.B. (Ed.) (1985) *Women and Science.* Lewes*:* The Falmer Press.
KELLER, E.F. (1985) *Reflections on Gender and Science.* New Haven: Yale University Press.
KELLY, A. (1981) *The Missing Half. Girls and Science Education.* Manchester University Press.
KELLY, A. (1985) The construction of masculine science. *British Journal of Sociology of Education*, 6, 133-154.
KELLY, A. (1986) The development of girls' and boys' attitudes to science - a longitudinal study. *European Journal of Science Education,* 8, 399-412.
KOBALLA Jr, T.R. (1992) Persuasion and attitude change in science education. *Journal of Research in Science Teaching*, 29, 63-80.
KOERTGE, N. (1994) Do feminists alienate women from the sciences? *The Education Digest*, 60, 4, 49-52.
LEVIN, T., SABAR, N. and LINMAN, Z. (1991) Achievement and attitudinal patterns of boys and girls in science. *Journal of Research in Science Teaching,* 29, 929-937.
LUCHINS, E.H. (1976) *Women mathematicians: a contemporary appraisal.* Paper presented at the annual meeting of the American Association for the Advancement of Science, Boston.
MATTERS, G. (1997) *Are Australian boys underachieving? An Analysis Using a Validity-Reliability Framework Based on the Work of Lee Cronbach and Pamela Moss*. Paper presented at 23rd Conference of the International Association for Educational Assessment in Durban South Africa.
MOIR, A., and JESSEL, D. (1989) *Brain Sex.* London: Michael Joseph.
MULLIS, I.V.S., and JENKINS, L.B. (1988) *The Science Report Card: Elements of Risk and Recovery.* (Trends and achievement based on the 1986 national assessment report No. 17-S-01). Princeton
MURPHY, P., and ELWOOD, J. (1997) *Gendered experiences, choices and achievement - Exploring the Links.* Paper presented at 23rd Conference of the International Association for Educational Assessment in Durban South Africa.
NATIONAL ASSESSMENT of EDUCATIONAL PROGRESS (NAEP). (1970) *Report 1, 1969-70 science: national results and illustrations and groups comparisons.* Denver: Education Commission of the States.
NATIONAL ASSESSMENT of EDUCATIONAL PROGRESS (NAEP) (1971) *Science: group results A.* Denver: Education Commission of the States.
NATIONAL ASSESSMENT of EDUCATIONAL PROGRESS (NAEP) (1977) *Science technical report: summary volume*. Denver: Education Commission of the States.
NITKO, A.J. (1983) *Educational Tests and Measurement: An Introduction.* New York: Harcourt Brace Jovanovich, Inc.
OFSTED/EOC (1996) *The gender divide: Performance differences between boys and girls at school. A report from the Office of Her Majesty's Chief Inspector of Schools and the Equal Opportunities Commission.* London: HMSO.
PARKER, L.H. and RENNIE, L.J. (1986) Sex stereotyped attitudes about science: can they be changed? *International Journal of Science Education,* 8, 173-183.
PARKER, R. (1985) Domestic Physics, *Times Educational Supplement,* 4 October.
PIBURN, L.H. and BAKER, D.R. (1993) If I were the teacher ... qualitative study of attitude towards science. *Science Education,* 77, 393-406.
REID, D.J., and RYLES, A.P. (1989) 'Balanced Sciences', the national curriculum and teachers' attitudes. *Research in Education,* 42-57.

ROBINSON-AWANA, P., KEHLE, J.T., and JENSON, R.W. (1986) But what about smart girls? Adolescent self esteem and sex-role perceptions as a function of academic achievement. *Journal of Educational Psychology*, 78, 179-83.
ROSSER, S.V. (1993) Female friendly science: Including women in curricular content and pedagogy in science, *The Journal of General Education*, 42, 3, 1991-220.
ROYAL SOCIETY and INSTITUTE of PHYSICS (1982) *Girls and Physics.* London: Royal Society and Institute of Physics.
RUSSELL, T., BLACK, P., HARLEN, W., JOHNSON, S., and PALACIO, D. (1988) *Science at age 11: A review of APU findings 1980-1984.* London: HMSO.
SCHOFIELD, B., BLACK, P., BELL, J.F., JOHNSON, S., MURPHY, P., QUALTER, A. and RUSSELL, T. (1988) *Science at age 13: A review of APU findings 1980-1984.* London: HMSO.
SJÖBERG, L. (1983) Interest, achievement and vocational choice. *European Journal of Science Education,* 5, 299-307.
SMAIL, B. (1985) An attempt to move mountains: the "Girls into Science and Technology" (GIST) project. *Journal of Curriculum Studies,* 17, 3, 351-354.
SMAIL, B. and KELLY, A. (1984) Sex differences in science among 11-year-old school children: affective. *Research in Science and Technological Education,* 2, 87-106.
SMAIL, B., WHYTE, J., and KELLY, A. (1982) Girls into science and technology: the first two years. *School Science Review,* 63, 620-630.
SMITHERS, A., and ROBINSON, P. (1994) *The Impact of Double Science.* The Engineering Council.

SOLOMON, J. (1992) The classroom discussion of science-based social issues presented on television: knowledge, attitudes and values. *International Journal of Science Education*, 14, 431-444.
SOLOMON, J. and HARRISON, K. (1991) Talking about science based issues: do boys and girls differ? *British Education Research Journal,* 17, 283-294.
STABLES, A. and HARVEY, T.J. (1986) Gender differences in attitudes to science for third year pupils: an argument for single-sex teaching in mixed schools. *Research in Science and Technological Education,* 4, 163-170.
STABLES, A. and STABLES, S. (1995). Gender differences in students approaches to A-level subject choices and perceptions of A-level subjects: a study of first-year A-level students in a tertiary college, *Educational Research*, 37, 1, 39-51.
SUNDAY TIMES (1994) Go-ahead girls leave experts with a mystery, 8 November.
SWATTON, P. (1994a) Aspects of pupils' ability to handle variables in different contexts. *Cambridge Journal of Education*, 24, 2, 213-229.
SWATTON, P. (1994b) Pupil's performance within the domain of data interpretation with particular reference to pattern recognition. *Research in Science and Technological Education,* 12, 129-142.
TABER, K.S. (1991) Gender Differences in Science Preferences on Starting Secondary School. *Research in Science and Technological Education,* 9, 2, 245-251.
TABER, K.S. (1991) Girl-friendly physics in the National Curriculum. *Physics Education,* 26, 221-225.
TABER, K.S. (1992) Science relatedness and gender appropriateness of careers: some pupils' perceptions. *Research in Science and Technological Education,* 10, 1, 105-115.
THE ROYAL SOCIETY (1982) *Science Education 11-18 in England and Wales*. London: The Royal Society.
TIMES EDUCATIONAL SUPPLEMENT (TES) (1991a) Test change threatens girls. 29 November.
TIMES EDUCATIONAL SUPPLEMENT (TES) (1991b) Exam changes 'will put girls off science', 12 December.
TIMES EDUCATIONAL SUPPLEMENT (TES) (1994) Girls urged to set boys an example. 11 November.
VERSEY, J. (1990) Taking action on gender issues in science education. *School Science Review,* 71, 9-14.
WARRINGTON, M. and YOUNGER, M. (1995) *Differential achievement of girls and boys at GCSE: Observations from the perspective of one school.* Homerton Research Report No. 4. Cambridge: Homerton College.
WATTS, D.M., and GILBERT, J.K. (1989) The new learning. *Studies in Science Education*, 16, 75-121.
WEINREICH-HASTE, H. (1981) The image of science, in A. Kelly (ed.) *The Missing half: girls and science education.* Manchester: Manchester University Press.

WELFORD, G., BELL, J. DAVEY, A., GAMBLE, R., GOTT, R. (1986) *Science in Schools. Age 15 Report No. 4.* London: APU.
WHYTE, J. (1986) *Girls into Science and Technology: The Story of a Project.* London: Routledge and Kegan Paul.
WOOLNOUGH, B.E. (1997). Motivating students or teaching pure science? *School Science Review,* 78, 67-72.
WYLIE, R. (1979) *The Self Concept: Theory and Research on Selected Topics, Vol. 2.* Lincoln: University of Nebraska Press.
ZIMMERER, L.K., and BENNETT, S.M. (1987) *Gender Differences on the California Statewide Assessment of Attitudes and Achievements in Science*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Multiple Choice Questions

Beller, M., and Gafni, N. (1995) *International perspectives on the schooling and learning achievement of girls and boys as revealed in the 1991 International Assessment of Educational Progress (IAEP).* Jersulem: National Institute of Testing and Evaluation.
Beller, M., and Gafni, N. (1996) *Can item format (multiple-choice vs, open-ended) account for gender differences in mathematics achievement?* Jerusalem: National Institute for Testing and Evaluation.
Bridgeman, B. (1993) *A comparison of open-ended and multiple-choice question formats for the quantitative section of the Graduate Record Examinations General* Tests (GRE Rep. No. 88-13P; ETS RR-91-35). Princeton, NJ: Educational Testing Service.
Bridgeman, B., and Roack, D.A. (1993) *Development and evaluation of computer-adminstered analytical questions for the Graduate Record Examinations General Test* (GRE Rep. No. 88-06P; ETS RR-92-49). Princeton, NJ: Educational Testing Service.
Dossey, J.A., Mullis, I.V., and Jones, C.O. (1993) *Can students do mathematical problem solving? Results form constructed-response questions in NAEP's 1992 mathematics assessment.* Princeton, NJ; Educational Testing Service, National Assessment of Education Progress.
Lawrence, I.M., Lyu, C.F., and Feigenbaum, M.D. (1995) *DIF data on free-response SAT I mathematical items* (ETS RR-95-22). Princeton, NJ: Educational Testing Service.